THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 272, No. 32, Issue of August 8, pp. 19827–19836, 1997
Printed in U.S.A.

Molecular Analysis of a Novel Winged Helix Protein, WIN

EXPRESSION PATTERN, DNA BINDING PROPERTY, AND ALTERNATIVE SPLICING WITHIN THE DNA BINDING DOMAIN*

(Received for publication, March 6, 1997, and in revised form, May 30, 1997)

Kwok-Ming Yao‡, Mi Sha, Zhijian Lu, and Gordon G. Wong

From the Genetics Institute, Inc., Cambridge, Massachusetts 02140

We have cloned a novel winged helix factor, WIN, from the rat insulinoma cell line, INS-1. Northern blot analysis demonstrated that WIN is highly expressed in a variety of insulinoma cell lines and rat embryonic pancreas and liver. In adults, WIN expression was detected in thymus, testis, lung, and several intestinal regions. We determined the DNA sequences bound *in vitro* by baculovirus-expressed WIN protein in a polymerase chain reaction-based selection procedure. WIN was found to bind with high affinity to the selected sequence 5′-AGATTGAGTA-3′, which is similar to the recently identified HNF-6 binding sequence 5′-DHWATTGAYT-WWD-3′ (where W = A or T, Y = T or C, H is not G, and D is not C). We have isolated human WIN cDNAs by library screening and 5′-rapid amplification of cDNA ends. Sequence analysis indicates that the carboxyl terminus of human WIN has been previously isolated as a putative phosphorylation substrate, MPM2-reactive phosphoprotein 2 (MPP2); WIN may be regulated by phosphorylation. Alignment of the rat and human WIN cDNAs and their comparison with mouse genomic sequence revealed that the WIN DNA binding domain is encoded by four exons, two of which (exons 4 and 6) are alternatively spliced to generate at least three classes of mRNA transcripts. These transcripts were shown by RNase protection assay to be differentially expressed in different tissues. Alternative splicing within the winged helix DNA binding domain might result in modulation of DNA binding specificity.

We are interested in the molecular basis of endocrine and exocrine pancreas formation. Gene expression studies suggest both pancreas compartments are derived from a band of endodermal cells in the foregut that comprises the pancreatic primordium. These specific endodermal cells can be identified prior to overt pancreas morphogenesis by their characteristic expression of Type II glucose transporter (Glut2) (1) and the homeobox gene *PDX-1* (2). A genetic deletion of the *PDX-1* gene results in an almost surgical deletion of the pancreas (3, 4). However, many additional transcription factors including HB9, Isl1, Neuro D/Beta 2, Nkx6.1, Pax6, and PTF1 are expressed in some cells of the pancreatic primordium and developing pancreas and may be important for complete pancreas development (5–10). Recently, analysis of Isl1- and Pax4-deficient embryos indicates that both transcription factors are required for endocrine islet cell formation (11, 12). Additional transcription factors may be involved.

The prototypical winged helix (WH)[1] factors (name based on the x-ray structure of HNF-3γ DNA-binding domain complexed to the transthyretin promoter) (13), *Drosophila melanogaster* Forkhead (Fkh) and rat HNF3 factors, are associated with the development of endodermal-derived tissues. Fkh mutants have an intestinal phenotype and HNF3 factors were initially isolated from the liver biochemically (14–16). The WH factors are likely to have a role in many endodermally derived organ including the pancreas.

Recent methods of degenerate PCR and low stringency hybridization have expanded the WH gene family (17–20). More than 80 members have been identified in different species. Their origins and functions have been reviewed extensively (21, 22). WH genes may have diverse roles evident by their expression beyond endodermal derivatives.

Functional diversity is evident in the wide spectrum of phenotypes associated with mutations of WH genes. HCM1 and FHL1 were isolated as suppressors of calmodulin and RNA polymerase III mutations, respectively, in yeast (23, 24). Genetic analysis revealed that *D. melanogaster croc* and *slp1,2* are required for proper segmentation in early embryogenesis (25, 26) and *Caenorhabditis elegans lin-31* is essential for normal vulva development (27). In rodents, natural mutations at the nude locus, which resulted in abnormal hair growth and thymus development, were shown to be due to the disruptions of the whn WH gene (28). The knockout phenotypes of at least three WH genes have been reported. The knockout of HNF3β led to defective node formation and the absence of notochord (29, 30). Brain abnormalities were detectable in knockout mice lacking expression of the neurally expressed BF-1 and BF-2 genes (31, 32). Moreover, loss of BF-2, which is also expressed in the stromal mesenchyme of the kidney, led to abnormal kidney morphogenesis (32).

In this paper, we describe the analysis of WH gene expression in a rat pancreatic endocrine cell line, INS-1, by RT-PCR and the subsequent isolation and characterization of a novel WH gene, named WIN. WIN has about 40% amino acid identity within the WH domain and was found to be highly expressed in different insulinoma cell lines and embryonic pancreas and liver. In adult tissues, WIN expression was high in testis and thymus and lower in lung and intestine. A histidine-tagged

* The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EBI Data Bank with accession number(s) U83112 and U83113.

‡ To whom correspondence should be addressed. Present address: Dept. of Biochemistry, University of Hong Kong, 3/F Li Shu Fan Bldg., 5 Sassoon Rd., Hong Kong. Tel.: 852-2819-9275; Fax: 852-2855-1254; E-mail:kmyao@hkusua.hku.hk.

[1] The abbreviations used are: WH, winged helix; PCR, polymerase chain reaction; RT, reverse transcription; SAAB, selection and amplification of binding sites; bp, base pair(s); kb, kilobase(s); RACE, rapid amplification of cDNA ends; ORF, open reading frame; RPA, RNase protection assay; MOI, multiplicity of infection; PAGE, polyacrylamide gel electrophoresis; EMSA, electrophoretic mobility shift assay.

WIN fusion protein was used to select the WIN binding sites *in vitro* by following a modified PCR-based selection and amplification of binding sites (SAAB) procedure. WIN has a unique binding specificity.

We isolated human WIN cDNAs and found that a region outside of the WH domain was previously isolated as a partial 3′ cDNA encoding MPM2-reactive phosphoprotein 2 (MPP2) (33). MPP2 was isolated by expression cloning with the MPM2 monoclonal antibody, which bound its phosphorylated epitopes. WIN may be regulated by phosphorylation at the carboxyl terminus. WIN function may also be regulated by differential splicing. Analysis of multiple human and rat WIN cDNAs indicated that differential splicing occurs within the WH DNA binding domain at regions important for directing DNA binding specificity (34). We demonstrated by RNase protection analysis that these unprecedented differential splicing events are regulated.

EXPERIMENTAL PROCEDURES

Standard molecular biology techniques used are described by Sambrook *et al.* (35). Total RNAs were extracted by the guanidium isothiocyanate method (36) and poly(A)$^+$ RNA prepared using the Promega Poly(A)Tract mRNA isolation system. PCR was done using Vent DNA polymerase (New England Biolabs, Inc.) unless specified otherwise. Sequencing was performed using the Sanger dideoxy chain termination method.

RT-PCR—The two sets of degenerate oligonucleotides, WH-1 (5′-AARCCHCCHTAWTCNTAYAT-3′) and WH-2 (5′-RTGYCKRATNGARTTCTGCCA-3′) were designed based on previous reports (18, 19). RT-PCR used the Perkin-Elmer RT-PCR kit with poly(A)$^+$ RNA from INS-1 cells as templates at an annealing temperature of 40 °C with random hexamers. The amplified DNA (~153 bp) was isolated and subcloned into pBluescriptII (Stratagene) using the TA cloning vector from Invitrogen.

Cloning of Rodent and Human WIN—A directional INS-1 cDNA library was constructed in plasmid vector, pdG4–5, using the Stratagene cDNA synthesis kit. The 3.0-kb rat WIN cDNA was isolated by screening one million colonies of this library using a 30-mer oligonucleotide (5′-GCCAGCCTGGCTTGGCAATCTGCTTAAAAT-3′). The human WIN cDNAs were isolated by screening human adenocarcinoma (Stratagene) and testis (CLONTECH) directional cDNA libraries using the rat WIN cDNA under high stringency conditions. 5′-RACE was performed using the Life Technologies, Inc. RACE kit with rat 18 days post coital pancreas total RNAs and human thymus total RNAs (CLONTECH) as templates. The longest 5′-RACE products were assembled with the rat and human partial cDNAs at unique *Eco*RV and *Bss*H1 site, respectively. The predicted ORF within the assembled 3.4-kb cDNA (WIN-1) was tested by coupled *in vitro* transcription/translation using the Promega TNT Coupled Reticulocyte Lysate System. The rat and human cDNA sequences were submitted to GenBank® under the accession numbers U83112 and U83113, respectively.

Expression Analysis (Northern Blots and RNase Protection Assay)—RNAs were electrophoresed on 1% agarose-formaldehyde gel and blotted onto nylon membrane (GeneScreen) and probed with ^{32}P-labeled WIN-1 cDNA. Blots were stripped and reprobed with rat γ-actin according to the GeneScreen manual. The CLONTECH mouse and human endocrine system Multiple Tissue Northern blots were probed with WIN-1 as described by the manufacturer using high stringency washing conditions. For RPA, WIN DNA spanning exons 4, 5, and 6 was amplified by PCR and subcloned into pBluescript II SK− as DNA template for RNA synthesis. After linearizing with *Eco*RI, ^{32}P-labeled antisense RNA probes (243 bases) were synthesized by *in vitro* transcription using T7 polymerase (Ambion Maxiscript kit) and gel-purified, and RPA was performed with total RNAs using the Ambion RPA kit. RPA using cyclophilin as probe was also carried out for RNA quantitation.

COS Transfection and Preparation of Nuclear Extracts—COS cells were transfected by the DEAE-dextran method (37). Two days after transfection, nuclear extracts were prepared from the cells according to Schreiber *et al.* (38).

WIN Protein Expression and Purification—The BAC-TO-BAC Baculovirus expression system (Life Technologies Inc.) was used to express the WIN protein. The WIN gene was generated by a two-step PCR procedure using three primers: Primer 1 (5′-CATCATCATGGAGACCATGACGATAAGATGAGAACCAGCCCCCGGCGG-3′), Primer 2 (5′-GTTGTTTGGATCCACCATGGGACACCATCACCATCATCATGGAGACGATGAC-3′), and Primer 3 (5′-GTTGTTTCTCGAGCTATCGCAGCTCACGGGATGAACTG-3′). PCR was performed first with Primers 1 and 3 for 10 cycles. The PCR product was purified using the Promega Wizard PCR Preps DNA Purification System, followed by PCR using Primers 2 and 3 for 20 additional cycles. The 5′-primers (Primers 1 and 2) led to the introduction of a *Bam*HI site, and then a sequence based on the Kozak rule for optimal protein expression in-frame with an initiating methionine with a glycine spacer followed by nucleotides coding for the $(HIS)_6$ tag and the enterokinase consensus sequence, followed by the first 21 nucleotides of the WIN gene. The 3′-primer (Primer 3) contains 24 nucleotides of 3′ WIN sequence and allowed the introduction of a *Xho*I site. The PCR product was digested by *Bam*HI and *Xho*I and ligated into identical sites of the donor plasmid pFASTBAC-1. The ligation product was transformed into DH10BAC *Escherichia coli* cell. The transformants were plated out in Luria agar plates containing kanamycin, gentamycin, tetracycline, blue-gal, and isopropyl-1-thio-β-D-galactopyranoside. Four white colonies were selected after 48 h of transformation. Mini DNA preparations were prepared, and the isolation of recombinant baculovirus DNA was confirmed by PCR. Transfection of Sf9 insect cells was by Cell-fectamine (Life Technologies, Inc.). The recombined virus was harvested after 7 days of transfection, and the virus stock was amplified by infecting Sf9 cells using low viral MOIs (1 MOI/cell). For WIN protein production, Sf9 cells were seeded to 90% confluence in two T175 flasks (Falcon), and the cells were infected with a high MOI (about 10 MOI/cell) from the viral stock. Infected cells were harvested after 96 h of infection and lysed in Tris buffer, pH 8.0, containing 0.5 M NaCl, 0.1% Nonidet P-40, 0.5 μg/ml leupeptin, 0.7 μg/ml pepstatin A, 0.2 μg/ml aprotinin, and 2 mM phenylmethylsulfonyl fluoride. Lysed cells were then sonicated briefly and centrifuged at 10,000 × *g* for 30 min. The supernatant was used for binding to an Ni column (Qiagen). The WIN protein was eluted out using 200 mM imidazole. WIN protein purification was confirmed by SDS-PAGE gels stained by Coomassie Blue.

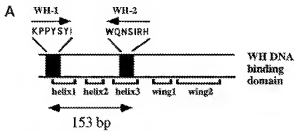
Electrophoretic Mobility Shift Assay—EMSA was conducted using the Bandshift kit from Pharmacia Biotech Inc. A typical DNA binding reaction contained ~2 ng of ^{32}P-labeled DNA and 2 μl of nuclear extract or purified WIN in 10 mM Tris-HCl, pH 7.5, 50 mM NaCl, 3 mM dithiothreitol, 5 mM $MgCl_2$, 0.05% Nonidet P-40, 10% glycerol, 1 μg poly (dI-dC), 0.5 μg/ml leupeptin, 0.7 μg/ml pepstatin A, 0.2 μg/ml aprotinin, and 2 mM phenylmethylsulfonyl fluoride at a total reaction volume of 20 μl. Both DNA binding and gel electrophoresis were carried out at 4 °C.

Selection and Amplification of Binding Sites—The DNA sequences recognized by WIN was determined using a modified SAAB procedure. The random oligonucleotide, 5′-CAGTGCTCTACAGGATCCGTCAC(N13)CGAAGCTTATCGATCCGAGCG-3′, and PCR primers (Primer 4, 5′-CGCTCGGATCGATAAGCTTCG-3′; Primer 5, 5′-CAGTGCTCTACAGGATCCGTGAC-3′) were designed according to Kunsch *et al.* (39). The random DNA pool for selection was generated by annealing of ^{32}P-labeled Primer 4 with the random oligonucleotide followed by Klenow extension. 500,000 cpm (~150 ng) of the labeled DNA was subjected to WIN binding and EMSA. In the first two rounds of selection, there was no discernible band shift, gel pieces above the unbound DNA were excised, and the DNA was eluted in TE (10 mM Tris, 1 mM EDTA, pH 8) with 50 mM NaCl. ~½ of the eluted DNA was amplified by PCR using Primers 4 and 5 for 30 cycles. After phenol/chloroform extraction, the amplified DNA was concentrated and washed in Microcon 100 concentrator (Amicom), followed by purification in a 12% native PAGE gel. The purified DNA was then radiolabeled by kinasing and subjected to subsequent round of WIN selection. After five rounds of WIN selection, the PCR-amplified DNA was digested with *Bam*HI and *Hin*dIII and subcloned into pBluescript II SK− for sequencing.

RESULTS

Isolation of Rat WIN—The insulinoma cell line INS-1 expresses many of the properties of isolated primary rat islet beta cells and is a ready source of material for gene expression analysis. We sought to characterize the WH genes expressed in INS-1 by PCR with two sets of degenerate oligonucleotides, WH-1 and WH-2, that span two conserved blocks of sequence homology within the WH DNA binding domain (Fig. 1*A*).

PCR products of about 150 bp were generated, subcloned, and sequenced (Fig. 1*B*). 35 clones were picked randomly and found to encode WH proteins; 51% of the clones showed identity to the HNF3γ DNA binding domain, 6% to the rat homolog of human FREAC, and 43% or 15 of 35 clones contained an identical novel WH sequence. Because the novel WH sequence



B Summary of PCR cloning

WH-like gene	Frequency	Expression
WIN (novel)	15/35 (43%)	this study
HNF3γ	18/35 (51%)	liver, gut
FREAC-4	2/35 (6%)	kidney, testis

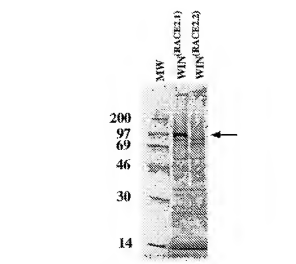


FIG. 1. **Cloning of WIN.** *A*, the proximal half of the winged helix domain containing most of the three helices (~153 bp) was amplified using degenerate primers, WH-1 and WH-2 (see "Experimental Procedures"), directed against the most highly conserved KPPYSYI and WQNSIRH regions. *B*, sequencing analysis of 35 PCR products encoding recognizable WH domains. 15 PCR products encode the novel WIN WH domain. The remaining 20 correspond to two previously cloned WH proteins, HNF3α and FREAC-4, and their known expression sites are also indicated. *C*, coupled *in vitro* transcription/translation of assembled rat WIN cDNAs. The near full-length cDNA (3.4-kb, WIN-1) was assembled using the longest 5′-RACE product, RACE2.1. This led to the generation of a 90-kDa polypeptide, agreeing with the predicted size of the 771-amino acid ORF (Fig. 2*A*). This polypeptide was not synthesized with a shorter cDNA assembled using RACE2.2, which starts at nucleotide 199 (Fig. 2*A*).

was cloned from INS-1 RNAs, we named the novel gene WIN (Winged helix from INS-1 cells).

Northern blot analysis of INS-1 RNAs indicated that the full-length cDNA gene for rat WIN should be about 3.5 kb (see Fig. 3*A*). We designed a 30-mer oligonucleotide from the novel WIN sequence and used it to screen a INS-1 cDNA library. A single clone with an insert of about 3 kb was isolated. DNA sequence analysis revealed an ORF of 651 amino acids containing the identified novel WH DNA binding domain, however without an initiating methionine. 5′-RACE with rat 18 dpc pancreas RNA generated a 900-bp fragment (RACE2.1) 5′ of the *Eco*RV site present in the 3-kb cDNA (Fig. 2*A*).

The 3-kb cDNA and RACE2.1 were assembled at the *Eco*RV site to give a 3.4-kb cDNA (WIN-1), which was completely sequenced (Fig. 2*A*). Conceptual translation revealed a 771-amino acid ORF that begins with two ATGs (at nucleotide 85). The absence of a purine in the −3 position of the first ATG would predict that the second ATG at nucleotide 88 is the initiating methionine. A similarly positioned methionine was found to be conserved in the human WIN cDNA sequence. Two in-frame stop codons are found 5′ to this ATG. WIN-1, when tested in a coupled *in vitro* transcription/translation reaction, yielded a polypeptide with a SDS-PAGE mobility of 90-kDa (Fig. 1*C*). A cDNA assembled using a shorter RACE fragment (RACE2.2) that starts at nucleotide 199 did not yield a translation product. The synthesis of WIN fusion protein of the predicted size using the baculovirus expression system also provides evidence that the predicted ORF was used *in vivo*.

WIN Is a Distant Relative of the WH Gene Family—WIN-1 was searched against GenBank® sequences. The only significant matches were gene sequences of the WH gene family and with MPP2 (see "Isolation of Human WIN"). From a comparison of the 10 most homologous WH genes, we found homology only in the WH DNA binding domain with no conservation of Regions II, III, and IV, previously identified as transcriptional activation domains in rodent HNF3s and other related WH proteins (21, 40). Both the alignment of the homologous WH domains against rat HNF3α (Fig. 2*B*) and the dendrogram analysis (Fig. 2*C*) indicate that WIN is distantly related to other WH proteins (less than 40% amino acid identity). The alignment also reveals the striking displacement of 12 amino acids in the center of Helix 3 of the WIN WH domain. The 36-bp DNA sequence corresponding to these 12 amino acids is absent from the original WIN PCR sequences.

We questioned whether this 36-bp DNA sequence would be evident in the genomic DNA sequence of WIN. Phage genomic DNAs for murine WIN were isolated, subcloned, and sequenced.[2] A comparison of mouse and rat sequences revealed the intron and exon structure described in Fig. 2*A*. The 36-bp sequence specific to the WH domain of the WIN gene is conserved in the mouse genomic WIN sequence and constitutes a single exon, exon 4. Moreover, RT-PCR analysis using primers flanking exon 4 and INS-1 poly(A)$^+$ RNA as templates indicated that both transcripts with and without exon 4 are expressed by INS-1 cells.

Analysis of WIN Expression by Northern Blots—WIN-1 was used as a probe for Northern analysis of RNAs from rodent and human cells and tissues (Fig. 3). Species specific RNA band patterns were observed: a 3.5-kb doublet and a faint 4.3-kb band in rat (Fig. 3, *A* and *B*); two equally intense 3.5-kb and 4.3-kb bands in mouse (Fig. 3, *A* and *C*) and a 4-kb band in human (Fig. 3*D*).

WIN expression was detected in all the rat (INS-1, B2, 38, and RIN56A) and murine (alphaTC1, betaTC1, and beta TC6) endocrine cell lines analyzed (Fig. 3*A*). PC12, a neuronal cell line, expressed a lower level of WIN. Rat RNAs prepared from e12, 14, 18, neonate and adult pancreas and livers were tested for WIN expression (Fig. 3*B*). Expression levels appeared to be high in the embryonic pancreas and liver but decreased to undetectable levels in the adult. The lack of detectable expression in HepG2 cells is consistent with the absence of expression in adult liver. However, expression of WIN could persist in islet endocrine cells and be diluted by its relatively low concentration in the adult pancreas. In adult tissues, high level WIN expression was apparent in testis and thymus (Fig. 3, *C* and *D*). A moderate level of WIN expression was also detected in lung and several intestinal regions (large intestine and duodenum; Fig. 6*B* and results not shown).

Expression of Functional WIN and SAAB Selection of WIN DNA Binding Sequences—The distant relationship of WIN to

[2] K.-M. Yao and G. G. Wong, unpublished data.

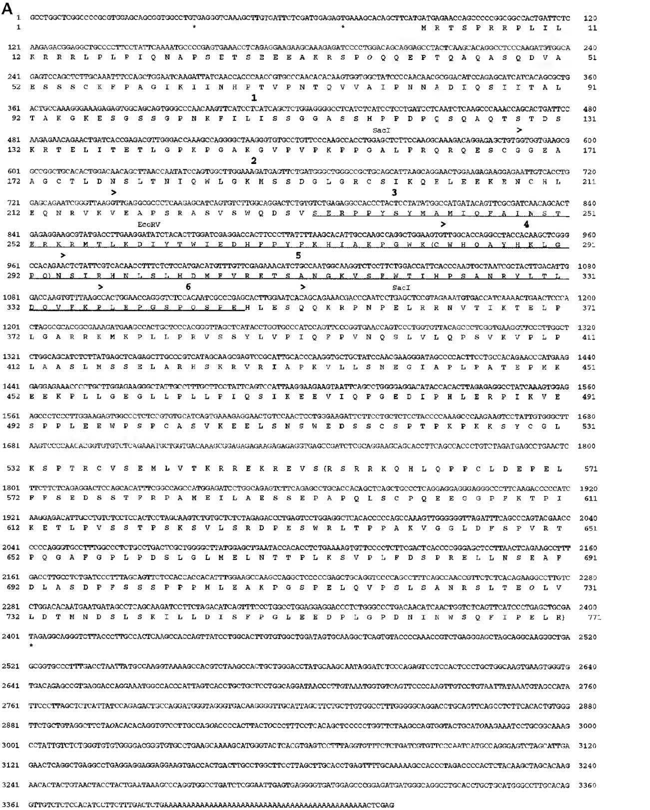


FIG. 2. **Sequence analysis of rat WIN.** *A*, sequence of the rat WIN cDNA and encoded protein. Positions of in-frame stop codons are denoted by *asterisks*. The WIN WH DNA binding domain identified by sequence comparison is *underlined*. Three restriction enzyme sites are indicated *above* the sequences. Also above the sequences are *arrowheads* and *numbers* that show the positions of the introns, predicted based on the comparison of the cDNA sequence against mouse genomic WIN sequence[2] and the corresponding assigned exons. *B* and *C*, comparison between WIN and 10 conserved WH proteins within the DNA binding domain using the Pileup comparison program (GCG). *B*, the sequences were aligned against rat HNF3α as a reference. The prefix letter in each sequence name denotes the sequence source: *r*, rat; *m*, mouse; *x*, frog; *d*, fruit fly; *h*, human; *y*, yeast. Within *parentheses* are the references of the sequences. *Dots* denote identical amino acids, and *dashes* represent gaps inserted in the sequences to optimize homology. The percentage of identity between any sequence against rHNF3α is indicated on the *right*. Predicted structures and divergent regions previously described within the WH domain (13) are shown *above* and *below* the sequence alignment. *C*, dendrogram analysis.

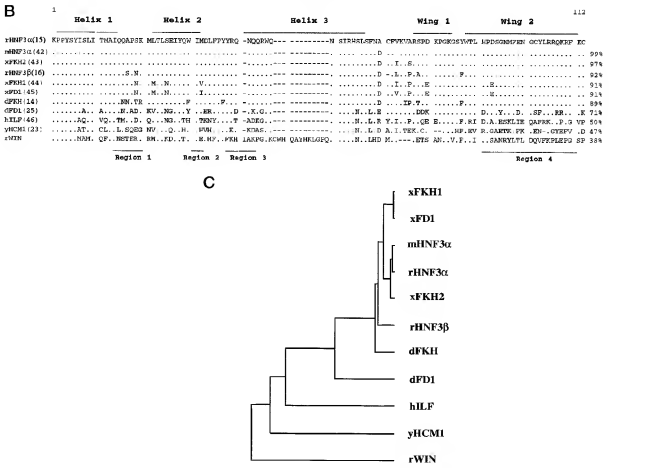


FIG. 2—*continued*

other WH proteins suggests that it may have a different DNA binding specificity. WIN-1, HNF3α, and HNF3γ cDNAs were heterologously expressed in COS-1 cells to generate nuclear extracts for DNA binding experiments. Nuclear extracts were prepared from transfected cells and tested for their ability to bind the known HNF3 binding sites in an EMSA. EMSA showed that nuclear extracts containing HNF3α and HNF3γ bound oligonucleotides corresponding to the HNF3 binding site TTR-S within the alpha transthyretin promoter (41) and the GluG2 site within the glucagon promoter (42), whereas binding was undetectable with the WIN extract (results not shown). In parallel transfection ^{35}S-labeled methionine was added to the COS-1 cell medium. Nuclear extracts analyzed by SDS-PAGE showed polypeptides corresponding in sizes to WIN, HNF3α, and HNF3γ were synthesized (results not shown).

We sought to determine the DNA sequence bound by heterologously expressed WIN protein in a PCR-based SAAB procedure. Because the COS-1 cell expression system yielded low amounts of WIN protein, which proved to be unsuitable for SAAB experiments, we chose to generate recombinant WIN protein using the high yield baculovirus system. The complete WIN ORF was inserted in-frame to an upstream Kozak sequence and a histidine tag in the baculovirus expression vector, pFASTBAC-1 (Life Technologies, Inc.). Transfected Sf9 cells were harvested, and total cellular extract was prepared and passed over a Ni-NTA affinity column. Partially purified WIN was recovered and analyzed by SDS-PAGE. Two specific protein bands were evident in the eluate; the predominant band of ~95 kDa was consistent with expression of the histidine-tagged full-length WIN protein and a second band of 50 kDa that was deduced to be a breakdown product (results not shown).

The recombinant WIN was used to select from a population of DNA oligonucleotides that consisted of a core of 13 randomized base pairs flanked by 5′ and 3′ PCR priming sites. After five rounds of selection and amplification, the prospective DNA binding sites were subcloned and sequenced. In later rounds of EMSA selection, two discrete mobility shift bands were observed, possibly due to the 95- and 50-kDa forms of the WIN protein. However, only the DNA oligonucleotides corresponding to the putative 95-kDa mobility shift were isolated and amplified. Similar EMSA analysis using a baculovirus cell extract without expressing the WIN protein and an unrelated histidine-tagged protein did not generate any detectable mobility shift.

26 cloned products from the final round of selection were sequenced. 15 of 26 clones sequenced were found to encode the identical sequence SAAB5–2, 5′-AGATTGAGTA-3′ (Fig. 4*A*). Radiolabeled oligonucleotide SAAB5–2 when combined with recombinant WIN protein showed the same two mobility shifts observed in the selection process (Fig. 4*B*, *lane 2*). The addition of 100 molar excess of unlabeled SAAB5–2 effectively displaced the radiolabeled oligonucleotides, suggesting that the WIN binding is specific (Fig. 4*B*, *lane 3*). We also tested three other SAAB-selected sequences for binding affinity by competing against radiolabeled SAAB5–2 in a competitive EMSA analysis (Fig. 4, *A* and *B*, *lanes 3, 13–15*). These three SAAB-selected sequences, SAAB5–12, SAAB5–1C, and SAAB5–13C, which displayed limited homology to SAAB5–2, could be bound by WIN when tested individually. They all displayed a moderate effect on SAAB5–2 binding, suggesting a lower binding affinity than SAAB5–2.

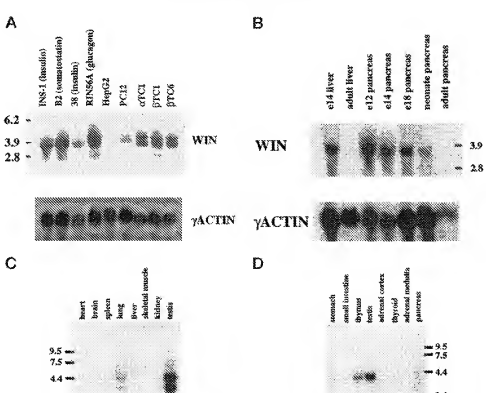


FIG. 3. **Northern blot analysis of WIN.** 20 μg of total RNA prepared from rodent insulinoma cell lines PC12 and HepG2 cells (*A*) and rat embryonic and adult liver and pancreas tissues (*B*) were electrophoresed on 1% agarose-formaldehyde gels, transferred onto nylon membranes (GeneScreen), and probed with ^{32}P-labeled rat WIN-1 cDNA. The Northern blots were reprobed with rat γ-actin to check RNA loading. Two CLONTECH Multiple Tissue Northern blots (mouse (*C*) and human endocrine system (*D*)) were similarly probed with the rat WIN-1 probe. 2 μg of poly(A)$^+$ RNA from each tissue was sampled. Different transcript patterns were observed for RNAs from different species (rat, 3.5-kb doublet and fainter 4.3 kb; mouse, 3.5 and 4.3 kb; human, 4 kb).

Next, we attempted to further test the specificity of WIN binding to SAAB5–2 by mutagenesis of the binding sequence (Fig. 4*B*, *lanes 4–8*). When the SAAB5–2 sequence was totally scrambled in mghSAAB5–2, its ability to compete against SAAB5–2 binding was eliminated. Selective mutations of the 5′ 2 bp, the 5′ 5 bp, the middle 5 bp, or 3′ 5 bp in mabSAAB5–2 (*lane 4*), mcdSAAB5–2 (*lane 5*), mjjSAAB5–2 (*lane 8*), and mefSAAB5–2 (*lane 6*), respectively, also significantly compromised their binding by WIN.

Sequence SAAB5–2 serves as an standard to evaluate additional prospective WIN binding sequences. SAAB5–2 matches 8 of 10 positions of the binding sequence, DHWATTGAYTWWD (Fig. 4*A*), of the recently characterized protein, HNF6 (43). HNF6 was demonstrated to bind HNF-3S.TTR at a lower affinity but not to the HNF-3#4 and HFH-1#3 binding sites. To compare the binding characteristics of WIN to HNF-6, we tested oligonucleotides comprising the binding sites for HNF6, HNF-3S.TTR, HNF3#4, and HFH-1#3 for their ability to competitively displace SAAB5–2 in EMSA (Fig. 4*B*, *lanes 9–12*). The extent of displacement suggests that WIN did bind to HNF6 and HNF3#4 oligonucleotides with greater affinity than HNF-3S.TTR and HFH-1#3, but it did so with lower affinity than SAAB5–2.

Isolation of Human WIN—A search of GenBank® revealed that the WIN-1 cDNA matched a human partial cDNA sequence encoding a 221 amino acid-protein termed MPP2 (33). MPP2 was isolated by expression cloning from a lymphoblast cell line cDNA library using the monclonal antibody MPM2 that bound a specific phosphorylated epitope. MPP2 had 76% identity at the amino acid level to the carboxyl-terminal 218 amino acids of rat WIN, which excludes the WH domain. This high degree of homology suggests that MPP2 might be the human homolog of WIN.

Directional cDNA libraries constructed from human pancreatic adenocarcinoma and from human testis were probed with a 605-bp *Sac*I fragment of WIN-1 that spans the WH domain (see Fig. 2*A*). Following high stringency hybridization and washing conditions, two clones were isolated from each library. All four clones were sequenced and found to have complete 3′ ends with sequences identical to the published MPP2 sequence and their 5′ sequences extending beyond MPP2 are highly homologous to the rat WIN-1 cDNA sequence, including the WH DNA binding domain. This observation strongly suggests that human WIN and MPP2 are identical genes.

The human WIN cDNAs extend to different lengths at the 5′ end and the longest cDNA of 3 kb is from the adenocarcinoma library. Comparison with the WIN-1 cDNA sequence indicated that the translation initiation codon was not reached. We synthesized the 5′ ends of WIN cDNAs by 5′-RACE from human thymus RNAs. The longest 5′-RACE product that contained the conserved initiating ATG and ~50 bp of 5′-untranslated leader sequence was assembled with the 3-kb human cDNA to generate a near full-length 3.34-kb human WIN cDNA that encoded a 764-amino acid ORF. Alignment of the human and rat WIN

A

SAAB5-2	AGATTGAGTA	(16/26)
SAAB5-12	TTAATTGGTCTC	(1/26)
SAAB5-1C	GTACTGCATTTAT	(1/26)
SAAB5-13C	GTTTAGTCTTTTCT	(1/26)
mabSAAB5-2	tcATTGAGTA	
mcdSAAB5-2	gtcgcGAGTA	
mefSAAB5-2	AGATTacacg	
mghSAAB5-2	gtcgcacacg	
mijSAAB5-2	AGcgcacGTA	
HNF-6	GATATTGATTTTT	
HNF-3S.TTR	GATTATTGACTT	
HNF-3#4	CAATGTTTGTTT	
HFH-1#3	AATTGTTTATTT	

B

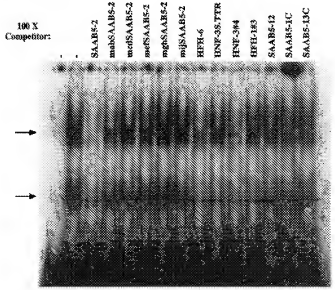


FIG. 4. **SAAB selection of DNA binding sequences.** *A*, sequences analyzed for WIN binding by EMSA. Double-stranded DNAs were assembled by annealing of complementary oligonucleotides with 5′- and 3′-flanking sequences (sense strand, 5′-TCGACGATCCGTCAC(N10–13)CGAAGCTTG-3′; antisense strand, 5′-TCGACAAGCTTCG(N10–13)GTCACGGATCG-3′). *SAAB5-x* represents four fifth round-selected sequences, and their abundance is shown in *parentheses*. The five mxxSAAB5–2 sequences are mutated versions of the SAAB5–2 sequences. The last four sequences (HNF-6, HNF-3S.TTR, HNF-3#4, and HFH-1#3) are natural or SAAB-selected sequences previously shown to be recognized by HNF3s (35, 44). *B*, competitive EMSA of binding sequences against the most abundant WIN SAAB-selected sequence, SAAB5–2. 2.5 ng of ^{32}P-labeled SAAB5–2 (~50,000 cpm) was subjected to WIN binding (~30 ng) in the absence (*lane 2*) and the presence of 100-fold molar excess of the indicated cold binding sequences (*lanes 3–15*). *Lane 1* represents the no protein control. The two DNA mobility shifts due to WIN binding are denoted by *arrows*.

amino acid sequences revealed stretches of extensive homology along the whole length of the protein (81% identity and 89% similarity; Fig. 5*A*). Seven of the nine potential phosphorylations sites identified in MPP2 (34) were also found in rat WIN.

Alternative Splicing within the DNA Binding Domain of WIN—When the rat and human WIN cDNAs including the 5′-RACE sequences were aligned, gaps of 36 and 45 bp became evident in the human cDNAs (Fig. 6*A*). These gaps correspond to exons 4 and 6, which fall within the WH DNA binding domain (see Fig. 2*A*). Exon 5 is present in all the isolated rat and human cDNAs and three classes of transcripts could be distinguished based on alternative splicing of exons 4 and 6. The rat INS-1 cDNA represents the Class a transcripts that contain all three exons, including exon 4 that is not present in other reported WH proteins. The two human pancreas cDNAs and thymus RACE products that lack exon 4 represent the Class b transcripts. Class c transcripts are represented by the two human testis cDNAs and thymus RACE products that lack both exons 4 and 6.

We attempted to determine the relative expression of the different WIN transcripts by RPA. A PCR-generated RPA probe spanning exons 4, 5, and 6 was used. Class a, b, and c transcripts would lead to the generation of protected bands of 210, 174, and 129 bp in length. Total RNAs from rat INS-1 cells, thymus, testis, and large intestine were analyzed (Fig. 6*B*). Both Class b and c transcripts are highly expressed in all the four tested RNAs, but their relative abundance varied. They were present at comparable levels in INS-1, thymus, and large intestine, but in testis Class c transcripts were present at a much higher level (Fig. 6*B*, *left panel*). Expression of Class a transcripts was detected at much lower levels in all four tested tissues (Fig. 6*B*, *right panel*). Class a transcripts in testis appeared to be expressed at a lower level relative to Class c transcripts.

This pattern of alternative splicing might have regulatory significance because exon 4 is within the region defined to be important for determining the DNA binding specificity and exon 6 within the Wing 2 region, which makes minor grove base-specific contacts (13, 34). We plan to test whether the different WIN protein isoforms encoded by the three transcript classes show different binding properties. The tissue-specific expression levels of the different WIN transcript Classes support the hypothesis proposing specific function with each alternately spliced transcript.

DISCUSSION

A novel WH protein, WIN, was isolated and found to be highly expressed in various insulinoma cell lines and early developing pancreas and liver by Northern analysis. In adult tissues, thymus and testis showed the highest level of expression, followed by lung and intestine at a lower level.

WIN is a divergent member of the WH gene family. Dendrogram analysis and pairwise comparisons within the WH domains show less than 40% amino acid identity between WIN and other WH members. No domains of homology exclusive of the WH domain could be identified. This divergence is also evident in the absence in WIN of a RK-rich sequence nuclear localization signal within W2 of HNF3β and in most if not all WH proteins (Ref. 44; see the alignment in Ref. 22).

We have found a high degree of homology between WIN and a human partial cDNA encoding an *in vitro* phosphorylated protein, MPP2. Using the rat WH domain DNA as a probe, we isolated near full-length human WIN cDNAs. The multiple human WIN cDNAs encode the reported MPP2 sequence at their 3′ ends, strongly suggesting that human MPP2 is human WIN.

Comparison of the rat and human WIN amino acid sequences allows us to define the putative boundaries of functional domains. The longest stretch of homology overlaps with the WH domain and, unlike other rat-human WH protein comparisons, extends beyond the normal carboxyl boundary of the WH domain. This could mean that the functional WIN DNA binding domain is about 100 amino acids longer than other WH proteins and the RK-rich areas within this extended portion may replace the function of similar basic sequences missing in the putative W2 region. Nine potential phosphorylation sites

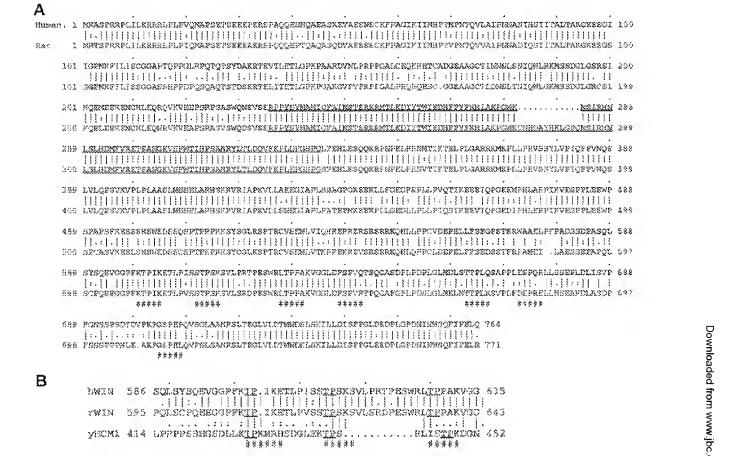


FIG. 5. **Sequence analysis of human WIN.** *A*, the human and rat full-length WIN amino acid sequences were aligned using the GAP comparison program (GCG). The WIN WH DNA binding domains are *underlined*. # indicates the positions of the seven putative phosphorylation sites identified in MPP2 that are conserved in rat WIN. *B*, alignment of the WIN sequences against the conserved putative phosphorylation sites of yeast HCM1. Alignment was performed using the Pileup comparison program (GCG). Four of the first six putative phosphorylation sites conserved between human and rat WINs (#) appear to be weakly conserved in yeast HCM1, and the conserved TP core amino acids are *underlined*.

with the central (T/S)P motif were predicted within human MPP2 based on comparison with peptide sequences of MPM2-reactive phosphorylated sites selected *in vitro*. The rat-human comparison also indicates that seven of the nine predicted putative phosphorylation sites are conserved.

Similar to WIN, yeast HCM1 also appears to lack the RK-rich sequence within the W2 region of the WH domain. HCM1 was originally isolated as a dosage-dependent suppressor of a calmodulin mutation cmd1–1 and appeared to enhance calmodulin function by an indirect mechanism (23). A visual comparison of HCM1 against rat and human WINs revealed a weak homology at their carboxyl termini (Fig. 5*B*). The central consensus TP amino acids corresponding to all four putative phosphorylation sites with a TP core appear to be conserved in HCM1. No similar homology was uncovered with other pairwise alignment with the WINs. In fact, when the complete yeast genomic sequence was searched using the rat WIN cDNA as query, HCM1 emerged as the WH protein with the best match. This together with the weak homology between WIN and HCM1 within the carboxyl-terminal of the proteins suggests that they might be more related members and that WIN can be tentatively placed within the Class 9b defined by Kaufmann and Knochel (22). It would also be interesting to test whether the four conserved putative phosphorylation sites are relevant to the regulation of WIN function.

As a first step toward understanding the DNA binding property of WIN *in vitro*, we prepared histidine-tagged full-length rat WIN protein for selecting DNA binding sites by the SAAB procedure. The purified WIN protein was very susceptible to proteolysis, and EMSA had to be performed at 4 °C in the presence of protease inhibitors. We found that thioredoxin and glutathione *S*-transferase fusion proteins with the carboxyl-terminal portion of WIN where the putative phosphorylation sites lie imparted instability to the fusion proteins. This finding agrees with the previous report that MPP2 was sensitive to proteolysis even in the presence of protease inhibitors and strong denaturants for MPP2 (33). The instability of WIN at room temperature might account for the lack of detectable binding to TTR-S using the WIN-transfected COS-1 nuclear extract.

After five rounds of SAAB selection, the 10-bp sequence, SAAB5–2, was highly enriched. SAAB5–2 is very similar to the recently reported HNF-6 binding site (43). We showed by competitive EMSA that WIN, like the HNF-6 binding activity, did bind the HNF-6 binding site and at a lower affinity the TTR-S site. However, WIN also bound to the HNF-3#4 site and to a lesser extent the HFH-1#3 site, to which HNF-6 did not bind. Thus, WIN and HNF-6 appeared to display similar but differ-

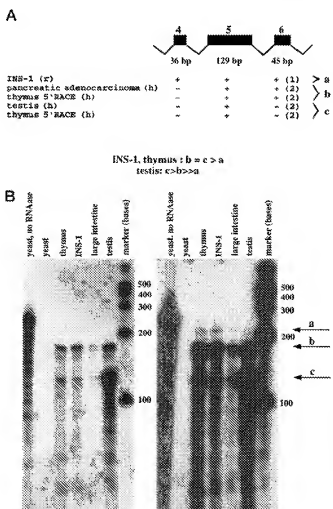


FIG. 6. **RPA of multiple WIN transcripts.** *A*, differential splicing pattern within the WIN DNA binding domain of multiple WIN cDNAs and summary of their relative abundance in different sources. Multiple rat (*r*) and human (*h*) cDNAs were isolated by library screening and 5′-RACE. The numbers of cDNAs isolated from different sources are indicated in *parentheses* on the *right*. They differ in their splicing pattern of exons 4 and 6. The *plus* and *minus signs* denote their presence in the different cDNAs. The rat INS-1 cDNA, which contains all three exons, corresponds to the Class a transcripts. Class b (missing exon) and c (missing exons 4 and 6) trancripts are represented by the different human pancreas, thymus, and testis cDNAs. *B*, RPA of WIN in different RNA sources. Total RNAs (50 μg) isolated from INS-1 cells, thymus, large intestine, and testis were subjected to RPA using a 243-base antisense WIN probe spanning exons 4, 5, and 6 (see "Experimental Procedures"). Equivalent loading was confirmed by RPA using cyclophilin as probe (results not shown). In the two control lanes (*yeast, no RNase* and *yeast*), total yeast RNAs were subjected to similar assay conditions in the presence and the absence of RNase. The Ambion Century template was used for the generation of the size markers. Class a, b, and c transcripts would lead to the generation of protected bands of 210, 174, and 129 bp in length, respectively (positions denoted by *arrows*). The *two panels* represent different exposure times: the *left panel* to show the higher relative abundance of Class c transcripts in testis and the *right panel* to show the presence of Class a transcripts.

ent DNA binding characteristics. It is very unlikely that WIN contributes to the HNF-6 binding activity because WIN mRNA was undetectable by Northern analysis in adult liver and HepG2 cells.

Another striking feature of the WIN gene was revealed by the comparison of genomic and cDNA sequences. The WH domain is interrupted by multiple introns, and exons 4 and 6 are alternatively spliced in cDNAs isolated from different tissue sources. Exon 4 is not conserved in any other reported WH members. Three Classes of transcripts (a, b, and c) arising from the splicing differences involving exons 4 and 6 were observed.

The relative abundance of these alternatively spliced transcripts was analyzed by RPA in different sources. Class a transcripts that contain exon 4 were expressed at a lower level than Class b and c transcripts. Expression of Class c transcripts was highly enriched in testis. This tissue-specific difference in transcript expression suggests that the splicing events may be regulated. The positions of exons 4 and 6 are interesting. Exon 4 lies within the region between H2 and H3, which was determined by Costa and co-workers (34) to be important for directing DNA binding specificity and exon 6 within W2, which was found to make the minor groove base-specific contacts (13). Taken together, these observations suggest that differential splicing within the WH domain may be of regulatory significance and the different protein isoforms generated may display diverse binding specificity. We have analyzed the DNA binding property of WIN encoded by the rat cDNA, which corresponds to a Class a transcript. It would be interesting to generate WIN isoforms corresponding to the more abundant Class b and c transcripts and test for differences in DNA binding property.

The expression of WIN in insulinoma cell lines and early developing pancreas (from e12 to neonate), when there is dramatic pancreas organogenesis, suggest that WIN may play a role in pancreas development. By RPA we have detected WIN expression in adult pancreas and islets. However, the expression of WIN in other tissues like e14 liver, thymus, testis, intestine, and fat from pregnant mothers (results not shown) suggests another hypothesis. Common among these tissues is the high content of mitotically active progenitor-like populations (hematopoietic in embryonic liver, T lymphocyte in thymus; germ cell in testis; intestinal in gut; adipocyte in fat from pregnant mothers; and exocrine and endocrine in embryonic pancreas). This, together with the observation that human MPP2/WIN is highly phosphorylated by M-phase kinases *in vitro*, suggests that WIN may be involved in the regulation of early progenitor cell growth. Consistent with this hypothesis, human WIN was recently found to be expressed in proliferating epithelial and mesenchymal cells of embryonic and adult tissues (45). Further RNA *in situ* hybridization and immunohistochemical analyses would be required to understand the detailed cellular level of WIN expression in the different tissues. In addition, we are currently testing the function of WIN by transgenic misexpression of WIN under the control of pancreas-specific promoters and ES-cell based gene knockout.

Acknowledgments—We thank Chris Miller for providing some of the RNAs for Northern analysis and Sarah Myers for helping in the library screening.

REFERENCES

1. Pang, K., Mukonoweshuro, C., and Wong, G. G. (1994) *Proc. Natl. Acad. Sci. U. S. A.* **91,** 9559–9563
2. Gus, Y., Montminy, M. R., Stein, R. W., Leonard, J., Gamer, L. W., Wright, C. V. E., and Teitelman, G. (1995) *Development* **121,** 11–18
3. Jonsson, J., Carlsson, L., Edlund, T., and Edlund, H. (1994) *Nature* **371,** 606–609
4. Offield, M. F., Jetton, T. L., Labosky, P. A., Ray, M., Stein, R. W., Magnuson, M. A., Hogan, B. L. M., and Wright, C. V. E. (1996) *Development* **122,** 983–995
5. Harrison, K. A., Druey, K. M., Deguchi, Y., Tuscano, J. M., and Kehrl, J. H. (1994) *J. Biol. Chem.* **269,** 19968–19975
6. Thor, S., Ericson, J., Brannstrom, T., and Edlund, T. (1991) *Neuron* **7,** 881–889
7. Naya, F. J., Strelkreet, C. M. M., and Tsai, M.-J. (1995) *Genes & Dev.* **9,** 1009–1019
8. Jensen, J., Serup, P., Karlsen, C., Nielsen, T. F., and Madsen, O. D. (1996) *J. Biol. Chem.* **271,** 18749–18758
9. Turque, N., Plaza, S., Radvanyi, F., Carriere, C., and Saule, S. (1994) *Mol. Endocrinol.* **8,** 928–938
10. Kragp, A., Knoffler, M., Frutiger, S., Hughes, G. J., Hagenbuchle, O., and Wellauer, P. K. (1996) *EMBO J.* **15,** 4317–4329
11. Ahlgren, U., Pfaff, S. L., Jessell, T. M., Edlund, T., and Edlund, H. (1997)

Nature **385,** 257–260
12. Sosa-Pineda, B., Chowdhury, K., Torres, M., Oliver, G., and Gruss, P. (1997) *Nature* **386,** 399–402
13. Clark, K. L., Halay, E. D., Lai, E., and Burley, S. K. (1993) *Nature* **364,** 412–420
14. Weigel, D., Jurgens, G., Kuttner, F., Seifert, E., and Jackle, H. (1989) *Cell* **57,** 645–658
15. Lai, E., Prezioso, V. R., Smith, E., Litvin, O., Costa, R. H., and Darnell, J. E., Jr. (1990) *Genes & Dev.* **4,** 1427–1436
16. Lai, E., Prezioso, V. R., Tao, W., Chen, W. S., and Darnell, Jr. J. E. (1991) *Genes & Dev.* **5,** 416–427
17. Hacker, U., Grossniklaus, U., Gehring, W. J., and Jackle, H. (1992) *Proc. Natl. Acad. Sci. U. S. A.* **89,** 8754–8758
18. Clevidence, D. E., Overdier, D. G., Tao, W., Qian, X., Pani, L., Lai, E., and Costa, R. H. (1993) *Proc. Natl. Acad. Sci. U. S. A.* **90,** 3948–3952
19. Kaestner, K. H., Lee, K.-H., Schlondorff, J., Hiemisch, H., Monaghan, A. P., and Schutz, G. (1993) *Proc. Natl. Acad. Sci. U. S. A.* **90,** 7628–7631
20. Pierrou, S., Hellqvist, M., Samuelsson, L., Enerback, S., and Carlsson, P. (1994) *EMBO J.* **13,** 5002–5012
21. Hromas, R., and Costa, R. (1995) *Crit. Rev. Oncol. Hematol.* **20,** 129–140
22. Kaufmann, E., and Knochel, W. (1996) *Mech. Dev.* **57,** 3–20
23. Zhu, G., Muller, E. G. D., Amacher, S. L., Northrop, J. L., and Davis, T. N. (1993) *Mol. Cell. Biol.* **13,** 1779–1787
24. Hermann-Le Denmat, S., Werner, M., Sentenac, A., and Thuriaux, P. (1994) *Mol. Cell. Biol.* **14,** 2905–2913
25. Hacker, U., Kaufmann, E., Hartmann, C., Jurgens, G., Knochel, W., and Jackle, H. (1995) *EMBO J.* **14,** 5306–5317
26. Grossniklaus, U., Pearson, R. K., and Gehring, W. J. (1992) *Genes & Dev.* **6,** 1030–1051
27. Miller, L. M., Gallegos, M. E., Morisseau, B. A., and Kim, S. K. (1993) *Genes & Dev.* **7,** 933–947
28. Nehls, M., Pfeifer, D., Schorpp, M., Hedrich, H., and Boehm, T. (1994) *Nature* **372,** 103–107
29. Ang, S.-L., and Rossant, J. (1994) *Cell* **78,** 561–574
30. Weinstein, D. C., Ruiz i Altaba, A., Chen, W. S., Hoodless, P., Prezioso, V. R., Jessell, T. M., and Darnell, J. E., Jr. (1994) *Cell* **78,** 575–588
31. Xuan, S., Baptista, C. A., Balas, G., Tao, W., Soares, V. C., and Lai, E. (1995) *Neuron* **14,** 1141–1152
32. Hatini, V., Huh, S. O., Herzlinger, D., Soares, V. C., and Lai, E. (1996) *Genes & Dev.* **10,** 1467–1478
33. Westendorf, J. M., Rao, P. N., and Gerace, L. (1994) *Proc. Natl. Acad. Sci. U. S. A.* **91,** 714–718
34. Overdier, D. G., Porcella, A., and Costa, R. H. (1994) *Mol. Cell. Biol.* **14,** 2755–2766
35. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
36. Davis, L. G., Dibner, M. D., and Battey, J. F. (1986) *Basic Methods in Molecular Biology*, pp. 129–135, Elsevier, New York
37. Danielsen, M., Northrop, J. P., and Ringold, G. M. (1986) *EMBO J.* **5,** 2513–2522
38. Schreiber, E., Matthias, P., Muller, M. M., and Schaffner, W. (1989) *Nucleic Acids Res.* **17,** 6419
39. Kunsch, C., Ruben, S. M., and Rosen, C. A. (1992) *Mol. Cell. Biol.* **12,** 4412–4421
40. Pani, L., Overdier, D. G., Porcella, A., Qian, X., Lai, E., and Costa, R. H. (1992) *Mol. Cell. Biol.* **12,** 3723–3732
41. Costa, R. H., Grayson, D. R., and Darnell, J. E., Jr. (1989) *Mol. Cell. Biol.* **9,** 1415–1425
42. Philippe, J., Morel, C., and Prezioso, V. (1994) *Mol. Cell. Biol.* **14,** 3514–3523
43. Samadani, U., and Costa, R. H. (1996) *Mol. Cell. Biol.* **16,** 6273–6284
44. Qian, X., and Costa, R. H. (1995) *Nucleic Acids Res.* **23,** 1184–1191
45. Ye, H., Kelly, T. F., Samadani, U., Lim, L., Rubio, S., Overdier, D. G., Roebuck, K. A., and Costa, R. H. (1997) *Mol. Cell. Biol.* **17,** 1626–1641